Statistical Inference - I

18K5S07

UNIT-I

Theory of Estimation - Definition, Parameter, statistic, Sampling distribution, Standard Error, Level of Significance, Utility of Standard Error, Hypothesis Definition of Null Hypothesis and Alternative Hypothesis Types-I and Type-II errors, one-tailed and two-tailed tests, critical region. Testing of hypothesis - General procedure.

UNIT-II

Point estimation - Properties of good estimator - consistency, unbiasedness efficiency and sufficiency. Cramer Rao inequality with Proof. Neymann factorisation theorem statement only simple Problems based on Binomial, Poisson normal and Exponential distribution.

UNIT-III

Methods of Estimation - Methods of Maximum likelihood estimation (MLE) and Methods of Moments - Simple Problems based on Binomial, Poisson Normal and Exponential distribution, Rao Blackwell theorem. Properties of MLE'S

UNIT-IV

Interval Estimation - Definition confidence Interval and confidence limits confidence limits based on normal distribution - confidence interval for single proportion and difference between proportion confidence Interval for single mean and difference of means - procedure and simple problems.

UNIT-V

Confidence Interval based on 't', 'F' and chi-square distribution confidence interval for single mean and difference of means. confidence Interval for variance and confidence interval for variance ratio- Procedure and simple problems.

UNIT-I

PARAMETER AND STATISTIC.

In order to avoid verbal confusion with the statistical constants of the population viz, mean(μ) variance σ^2 etc. which are usually reffered to as parameters, statistical measures computed from the sample observations alone eg. mean($\bar{x}$), variance(σ^2), etc.

In practice parameter values are not Known and the estimates based on the sample values are generally used. Thus, statistic which may be regarded as an estimate of parameter, obtained from the sample, is a function of the sample values only.

Sampling Distribution of a Statistic.

If we draw a sample of size n from a given finite population of size N, then the total number of possible samples is :

$$N C_n = \frac{N!}{n!\,(N-n)!} = K,\ (say)$$

For each of these K samples we can compute some statistic $t = t(x_1, x_2, \ldots x_n)$ in particular the mean $\bar{x}$, the variance s^2 etc, as given below.

Sample Number	Statistic		
	t	$\bar{x}$	s^2
1	t_1	$\bar{x}_1$	s_1^2
2	t_2	$\bar{x}_2$	s_2^2
3	t_3	$\bar{x}_3$	s_3^2
⋮	⋮	⋮	⋮
K	t_K	$\bar{x}_K$	s_K^2

The set of the values of the statistic so obtained, one for each sample, constitutes what is called the sampling distribution of the statistic (For example the values $t_1, t_2, t_3, \ldots, t_k$ determine the sampling distribution of the statistic t. In other words, statistic 't' may be regarded as a random variable which can take the values $t_1, t_2, t_3 \ldots t_k$ and we can compute the various statistical constants like mean, variance, skewness, kurtosis, etc for its distribution.

Standard Error:

The standard deviation of the sampling distribution of a statistic is known as its standard Error, abbreviated as S.E. The standard errors of some of the well-known statistics, for large samples, are given below, where n is the sample size, σ^2 the population variance, and p the population proportion and $Q = 1 - P$; n_1 and n_2 represent the sizes of two independent random samples, respectively drawn from the given population(s).

S.No	Statistics	Standard Error
1)	Sample Mean : $\bar{x}$	$\sigma / \sqrt{n}$
2)	observed sample Proportion 'p'	$\sqrt{PQ/n}$
3)	Sample s.d. : s	$\sqrt{\sigma^2 / 2n}$

4) Sample variation : s^2	$\sigma^2 \sqrt{2/n}$
5) Sample quartiles	$1.36263\, \sigma/\sqrt{n}$
6) Sample median	$1.25331\, \sigma/\sqrt{n}$
7) Sample correlation coefficient (r)	$(1-\rho^2)/\sqrt{n}$, ρ being the population correlation coefficient.
8) Sample moment : μ_3	$\sigma^3 \sqrt{96/n}$
9) Sample moment : μ_4	$\sigma^4 \sqrt{96/n}$
10) Sample coefficient of variation (v)	$\frac{v}{\sqrt{2n}} \sqrt{1 + \frac{2v^3}{10^4}} = \frac{v}{\sqrt{2n}}$
11) Difference of two sample means : $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_3^2}{n_2}}$
12) Difference of two sample s.d.'s : $(s_1 - s_2)$	$\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_3^2}{2n_2}}$
13) Difference of two sample proportions : $(P_1 - P_2)$	$\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

Utility of Standard Error.

S.E. Plays a very important role in the large sample theory and forms the basis of the testing of hypothesis. If t is any statistic, then for large samples: $z = \frac{t - E(t)}{\sqrt{V(t)}} \sim N(0,1)$

$$z = \frac{t - E(t)}{S.E.(t)} \sim N(0,1), \text{ for large}$$

samples.

Tests of significance:

A very important aspect of the sampling theory is the study of the tests of significance which enable us to decide on the basis of the sample results, if

(i) the derivation between the observed sample statistics and the hypothetical parameter value, or

(ii) the deviation between two independent sample statistics; is significant or might be attributed to chance or the fluctuations of sampling.

Since for large n, almost all the distribution, e.g., Binomial, Poisson, Negative binomial, Hyper geometric, t, F, chi-square, can be approximated very closely by a normal probability curve, we use the Normal test of significance for large samples. Some of well-known tests of significance for studying such differences for small samples are t-test, F-test and Fisher's z-transformation.

Null and Alternative Hypotheses:

The technique of randomisation used for the selection of sample units makes the test of significance valid for us. For applying the test of significance we first set up a hypothesis — a definite statement about the population parameter. Such a hypothesis, which is usually a hypothesis of no difference, is called null hypothesis and is usually denoted by H_0. According to prof. R.A. Fisher null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true.

For example, in case of a single statistic, H_0 will be that the sample statistics does not differ significantly from the hypothetical parameter value and in case of two statistics, H_0 will be that the sample statistics do not differ significantly.

~~Having~~ Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H_1. For example, If we want to test the null hypothesis that the population has a specified mean μ_0 (say), i.e, $H_0: \mu = \mu_0$ then the alternative hypothesis could be:

(i) $H_1: \mu \neq \mu_0$ (i.e, $\mu > \mu_0$ or $\mu < \mu_0$)

(ii) $H_1: \mu > \mu_0$

(iii) $H_1: \mu < \mu_0$

The alternative hypothesis in (i) is known as a two-tailed alternative and the alternatives in (ii) and (iii) are known as right-tailed and left-tailed alternatives respectively.

Errors in sampling:

The main objective in sampling theory is to draw valid interferences about the population parameters on the basis of the sample results. In practice we decide to accept or reject the lot after examining a sample from it. As such we are liable to commit the following two types of errors:

Type I Error: Reject H_0 when it is true.

Type 2 Error: Accept H_0 when it is wrong, i.e. accept H_0 when H_1 is true.

If we write P(Reject H_0 when it is true)

$$= P(\text{Reject } H_0 \mid H_0) = \alpha$$

and

P(Accept H_0 when it is wrong $= P(\text{Accept } H_0 \mid H_1) = \beta$

then α and β are called the sizes of type I error and type II error, respectively

In practice, type I error amounts to rejecting a lot when it is good and type II error may be regarded as accepting the lot when it is bad.

Thus

$$\left.\begin{aligned} &P(\text{Reject a lot when it is good}) = \alpha \\ &\text{and} \\ &P(\text{Accept a lot when it is bad}) = \beta \end{aligned}\right\} \quad 1a$$

where α and β are referred to as producer's risk and consumer's ~~right~~ risk respectively.

Critical region and Level of Significance.

A region (corresponding to a statistic t) in the sample space S which amounts to rejection of H_0 is termed as critical region of rejection. If ω is the critical region of and if $t = t(x_1, x_2, \ldots, x_n)$ is the value of the statistic based on a random sample of size n, then

$$P(t \in \omega \mid H_0) = \alpha, \quad P(t \in \bar{\omega} \mid H_1) = \beta \longrightarrow (1)b$$

where $\bar{\omega}$, the complementary set of ω, is called the acceptance region.

We have $\omega \cup \bar{\omega} = S$ and $\omega \cap \omega = \phi$

The probability 'α' that a random value of the statistics t belongs to the critical region is known as the level of significance. In other words, level of significance is the size of the type I error (or the maximum producer's risk) the levels of significance usually employed in testing of hypothesis are 5% and 1%. The level of significance is always fixed in advance before collecting the sample information.

One tailed and two tailed Tests

A test of any statistical hypothesis where the alternative hypothesis is one tailed (right tailed or left tailed) is called a one tailed test

For example:

a test for testing the mean of a population $H_0 : \mu = \mu_0$ against the alternative hypothesis

$H_1 : \mu > \mu_0$ (Right tailed) or $H_1 : \mu < \mu_0$ (left - tailed) is a single tailed test.

In the right - tailed test ($H_1 : \mu > \mu_0$), the critical region lies entirely in the right tail of the sampling distribution of $\bar{x}$, while for the left - tailed test ($H_1 : \mu < \mu_0$), the critical region is entirely in the left tail of the distribution.

A test of statistical hypothesis where the alternative hypothesis is two - tailed such as $H_0 : \mu \neq \mu_0$ against the alternative hypothesis $H_1 : \mu \neq \mu_0$ ($\mu > \mu_0$) and ($\mu < \mu_0$) is known as two tailed test.

For example

Suppose that there are two population brands of bulbs one manufactured by standard process (with mean life μ_1) and the other manufactured by some new techniques (with mean life

If we want to test if the bulbs differ significantly, then our null hypothesis is $H_0 : \mu_1 = \mu_2$ and alternative will be $H_1 : \mu_1 \neq \mu_2$, thus giving us two-tailed test.

Critical value or Significant values.

The value of test statistic which separates the critical (or rejection) region and the acceptance region is called the critical value or significant value. It depends upon:

(i) The level of significance used, and

(ii) The alternative hypothesis, whether it is two-tailed or single-tailed.

As has been pointed out earlier, for large samples the standardised variable corresponding to the statistic t, viz,

$$z = \frac{t - E(t)}{S.E(t)} \sim N(0,1) \longrightarrow (1)$$

asymptotically as $n \to \infty$. The Value of Z given by (1) under the null hypothesis is known as test statistic. The critical Value of the test statistic at level of significance α for a two-tailed test is given by z_α, where z_α is determined by the equation:

$$P(|Z| > z_\alpha) = \alpha \longrightarrow (2)$$

i.e) z_α is the value so that the total area of the critical region on both tails is α. Since normal probability curve is a symmetrical curve, from (2), we get

$$P(Z > z_\alpha) + P(Z < -z_\alpha) = \alpha$$

$$\Rightarrow P(Z > z_\alpha) + P(Z > z_\alpha) = \alpha \quad \text{[By symmetry]}$$

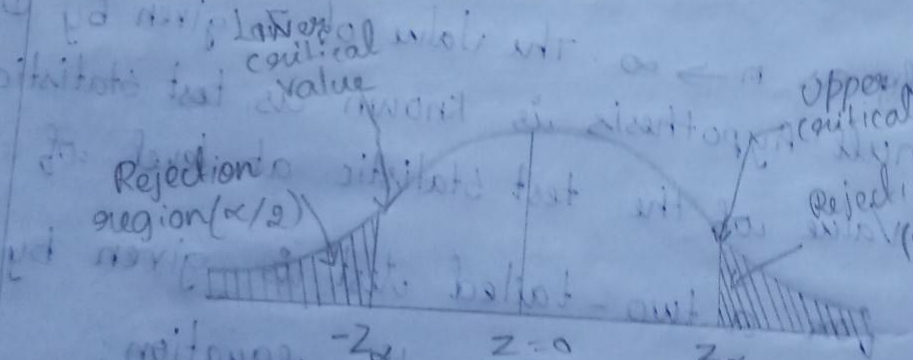
$$\Rightarrow 2P(Z > z_\alpha) = \alpha$$

$$\Rightarrow P(Z > z_\alpha) = \alpha/2$$

In other words, the area of each tail is $\alpha/2$. Thus z_α is the value such that area to the right of z_α is $\alpha/2$ and to the left of $(-z_\alpha)$ is $\alpha/2$, as shown in the following diagram.

TWO - TAILED TEST

(Level of significance 'α')



In case of single-tail alternative, the critical value z_α is determined so that total area to the right of it (for right-tailed test) is α and for left-tailed test the total area to the left $(-z_\alpha)$ is α (See diagrams below), i.e.,

Procedure for testing of Hypothesis:

1. Null Hypothesis: Set up the Null hypothesis H_0.

2. Alternative Hypothesis: Set up the Alternative Hypothesis H_1. This will enable us to decide whether we have to use a single-tailed (right or left) test or two-tailed test.

3. Level of significance: Choose the appropriate level of significance (α) depending on the reliability of the estimate and permissible risk. This is to be decided before sample is drawn i.e, α is fixed in advance.

4) Test statistic (or Test Criterion) Compute the test statistic:

$$Z = \frac{t - E(t)}{S.E(t)}, \text{under } H_0$$

5) Conclusion: We compare the computed value of Z in step 4 with the significant value (tabulated value) Z_α at the given level of significance, 'α'

If $|Z| < Z_\alpha$, ie, if the calculated value of Z (in modoulus value) is less than Z_α we say its is not significant. By this we mean that the difference $t - E(t)$ is just due to fluctuations of sampling and the sample data do not provide us sufficient evidence against the null hypothesis which may, therefore, be accepted.

If $|Z| > Z_\alpha$ ie if the computed value of test statistic is greater than the critical or significant value, then we say that it is significant and the null hypothesis is rejected at level of significance α, with confidence coefficient $(1-\alpha)$.

Test of Significance for large Samples:

In this section, we will discuss the tests of significance when samples are large. We have seen that for large values of n, the number of trials, almost all the distributions eg, binomial, Poisson, negative binomial, etc., are very closely approximated by normal distribution. thus in this case we apply the normal test, which is based upon the fundamental property (area property) of the normal probability curve

If $X \sim N(\mu, \sigma^2)$, then

$$z = \frac{X-\mu}{\sigma} = \frac{X - E(X)}{\sqrt{V(X)}} \sim N(0,1).$$

UNIT - 2

Definitions

Any function of the random sample $x_1, x_2, \ldots x_n$ that are being observed. Say $T_n(x_1, x_2, \ldots x_n)$ is called a statistic. clearly, a statistic is a random variable. If it is used to estimate an unknown parameter θ of the distribution, It is called an estimator. A particular value of the estimator, say $T_n(x_1, x_2, \ldots x_n)$ is called an estimate of θ.

Characteristics of Estimators:

The fg are some of the criteria that should be satisfied by a good estimator.
(i) unbiasedness (ii) consistency (iii) Efficiency, and (iv) sufficiency
we shall now, briefly explain these terms one by one.

unbiasedness:

Definitions:-

An estimator $T_n = T(x_1, x_2, \ldots, x_n)$ is said to be an unbiased estimator of $\gamma(\theta)$ if

$$E(T_n) = \gamma(\theta), \text{ for all } \theta \in \Theta$$

We have seen in chapter 13 that in sampling from a population with mean μ and variance σ^2, $E(\bar{x}) = \mu$ and $E(s^2) \neq \sigma^2$ but $E(S)^2 = \sigma^2$, Hence there is reason to prefer.

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$, to the sample Variance

$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

Remark:

If $E(T_n) > \theta$, T_n is said to be positively biased and if $E(T_n) < \theta$, it is said to be negatively biased, the amount of bias $b(\theta)$ being given by.

$$b(\theta) = E(T_n) - \gamma(\theta),\ \theta \in \Theta$$

Consistency:

Definitions:-

An estimator $T_n = T(x_1, x_2, \cdots x_n)$ based on a random sample of size n, is said to be consistent estimator of $\gamma(\theta)$, $\theta \in \Theta$, the parameter space, if T_n converges to $\gamma(\theta)$ in Probability, i.e., is $T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$. In other words. T_n is a consistent estimator of $\gamma(\theta)$ if for every $\varepsilon > 0$, $\eta > 0$, there exists a positive integer $n \geq m(\varepsilon, \eta)$ such that.

$$P\{|T_n - \gamma(\theta)| < \epsilon\} \to 1 \text{ as } n \to \infty \Rightarrow P\{|T_n - \gamma(\theta)| < \epsilon\} > 1-\eta;\ \forall n \geq m, \ldots$$

where m is some very large value of n.

Remark:

1, If $x_1, x_2, \dots x_n$ is a random sample from population with finite mean $E x_i = \mu < \infty$, then by Khinchine's weak law of large numbers (W.L.L.N) we have

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{P} E(x_i) = \mu \text{ , as } n \to \infty$$

Hence sample mean $(\bar{x}_n)$ is always a consistent estimator of the population mean(μ)

2, Obviously consistency is a property concerning the behaviour of an estimator for indefinitely large values of the sample size n, i.e., as $n \to \infty$. Nothing is regarded of its behaviour for finite n.

Moreover, if there exists a consistent estimator, say. T_n of $\gamma(\theta)$, then indefinitely many such estimator can be constructed. eg.,

$$T_n' = \left(\frac{n-a}{n-b}\right) T_n = \left[\frac{1-(a/n)}{1-(b/n)}\right] T_n \to T_n \xrightarrow{P} \gamma(\theta) \text{ , as } n \to 0$$

and hence, for different values of a and b. T_n' is also consistent for $\gamma(\theta)$.

Invariance Property of Consistent Estimators:-

Theorem:

In T_n is a consistent estimator of $\gamma(\theta)$ and $\psi\{\gamma(\theta)\}$ is a continuous function of $\gamma(\theta)$, then $\psi(T_n)$ is a consistent ~~of~~ estimator of $\psi\{\gamma(\theta)\}$.

Proof:

Since T_n is a consistent estimator of $\gamma(\theta)$ and $T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$, i.e., for every $\epsilon > 0$, $\eta > 0$, $\exists$ a positive integer $n \geq m(\epsilon, \eta)$ such that

$P\{|T_n-\gamma(\theta)|<\epsilon\} > 1-\eta$, $\forall n \geq m$.

Since $\psi(\cdot)$ is a continuous function, for every $\epsilon>0$, however small, $\exists$ a positive number ϵ_1 such that $|\psi(T_n)-\psi\{\gamma(\theta)\}|<\epsilon_1$ whenever $|T_n-\gamma(\theta)|<\epsilon$ i.e.,

$|T_n-\gamma(\theta)|<\epsilon \Rightarrow |\psi(T_n)-\psi\{\gamma(\theta)\}|\ \epsilon_1$ for two events A and B, if $A \Rightarrow B$, then

$$A \leq B \Rightarrow P(A) \leq P(B)$$
$$P(B) \geq P(A)$$

From (2) and (3) we get

$$P[|\psi(T_n)-\psi\{\gamma(\theta)\}|<\epsilon_1] \geq P[|T_n-\gamma(\theta)|<\epsilon]$$

$$P[|\psi(T_n)-\psi\{\gamma(\theta)\}|<\epsilon_1] \geq 1-\eta;\ \forall n \geq m$$

$\psi(T_n) \xrightarrow{P} \psi\{\gamma(\theta)\}$, as $n\to\infty$ or $\psi(T_n)$ is a consistent estimator of $\gamma(\theta)$.

Sufficient conditions for consistency:

Theorem: 2

Let $\{T_n\}$ be a sequence of estimators such that for all $\theta \in \Theta$

(i) $E_\theta(T_n) \to \gamma(\theta)$. $n\to\infty$ and

(ii) $\mathrm{var}_\theta(T_n) \to 0$, as $n\to\infty$

Then T_n is a consistent estimator of $\gamma(\theta)$.

Proof:

We have to prove that T_n is a consistent estimator of $\gamma(\theta)$.

$$T_n \xrightarrow{P} \gamma(\theta),\ \text{as } n\to\infty$$

i.e., $P[|T_n - \gamma(\theta)| < \epsilon] > 1 - \eta \; ; \; \forall \, n \geq m(\epsilon, \eta) \rightarrow ③$

where ϵ and η are arbitrarily small positive numbers and m is some large value of n.

Applying chebychev's inequality to the statistic T_n, we get.

$$P[|T_n - E_\theta(T_n)| \leq \delta] \geq 1 - \frac{Var_\theta(T_n)}{\delta^2} \rightarrow ④$$

We have,

$$|T_n - \gamma(\theta)| = |T_n - E(T_n) + E(T_n) - \gamma(\theta)| \leq |T_n - E_\theta(T_n)| + |E_\theta(T_n) - \gamma(\theta)| \rightarrow ⑤$$

Now,

$$|T_n - E_\theta(T_n)| \leq \delta \Rightarrow |T_n - \gamma(\theta)| \leq \delta + |E_\theta(T_n) - \gamma(\theta)| \rightarrow ⑥$$

Hence on using (* **) of theorem ① weget,

$$P\{|T_n - \gamma(\theta)| \leq \delta + |E_\theta(T_n) - \gamma(\theta)|\} \geq P\{|T_n - E_\theta(T_n)| \leq \delta\} \geq 1 - \frac{Var_\theta(T_n)}{\delta^2} \rightarrow ⑦$$

We are given:

$$E_\theta(T_n) \rightarrow \gamma(\theta) \; \forall \, \theta \in \Theta \text{ as } n \rightarrow \infty$$

Hence, for every $\delta_1 > 0$, $\exists$ a positive integer $n \geq n_0(\delta_1)$ such that

$$|E_\theta(T_n) - \gamma(\theta)| \leq \delta_1 \; \forall \, n \geq n_0(\delta_1) \rightarrow ⑧$$

Also $Var_\theta(T_n) \rightarrow 0$ as $n \rightarrow \infty$, (Given):

$$\frac{Var_\theta(T_n)}{\delta^2} \leq \eta, \; \forall \, n \geq n_0'(\eta) \rightarrow ⑨$$

where η is orbitarily small positive number

substituting from ⑧ and ⑨ in ⑦ we get

$$P[|T_n - \gamma(\theta)| \leq \delta + \delta_1] \geq 1 - \eta \; ; \; n \geq m(\delta_1, \eta)$$

$\Rightarrow P[|T_n - \gamma(\theta)| \leq \epsilon] \geq 1-\eta$; $n \geq m$.

where $m = \max(n_0, n_0')$ and $\epsilon = \delta + \delta_1 > 0$

$\Rightarrow T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$

T_n is a consistent estimator of $\gamma(\theta)$.

Efficient Estimators:

Efficiency, Even if we confine ourselves to unbiased estimates, there will, in general exist more than one consistent estimator of a parameter For example, in sampling from a normal population $N(\mu, \sigma^2)$ when σ^2 is known, sample mean $\bar{x}$ is an unbiased and consistent of μ [c.f. Example 17.5a]

From symmetry it follows immediately that sample median (Md) is an unbiased estimate of μ, which is same as the population median. Also for large n,

$$V(Md) = \frac{1}{4nf_1^2}$$

Here,

f_1 = Median ordinate of the parent distribution

= Modal ordinate of the parent distribution.

$$= \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-(x-\mu)^2/2\sigma^2\right\}\right]_{x=\mu} = \frac{1}{\sigma\sqrt{2\pi}}$$

$$\therefore V(Md) = \frac{1}{4n} \cdot 2\pi\sigma^2 = \frac{\pi\sigma^2}{2n}$$

Since $\left.\begin{array}{l} E(Md) = \mu \\ V(Md) \to 0 \end{array}\right\}$, as $n \to \infty$

median is also an unbiased and consistent estimator of μ.

Thus, there is a necessity of some further criterion which will enable us to choose between the estimators with the common property of consistency. Such a criterion which is based on the variances of the sampling distribution of estimators is usually known as efficiency. If, of the two consistent estimators T_1, T_2, of a certain Parameter θ, we have

$$V(T_1) < V(T_2), \text{ for all } n.$$

Then T_1 is more efficient than T_2 for all samples sizes.

We have seen above.

For all n, $V(\bar{x}) = \frac{\sigma^2}{n}$ and for large n,

$$V(Md) = \frac{\pi\sigma^2}{2n} = 1.57\frac{\sigma^2}{n}$$

Since $V(\bar{x}) < V(Md)$, we conclude that for normal distribution, sample mean is more efficient estimator for μ than the sample median, for large samples at least.

Efficiency:

Definitions:

If T_1 is the most efficient estimator with variance V_1 and T_2 is any other estimator with variance V_2, then efficiency E of T_2 is defined as:

$$E = V_1/V_2$$

obviously E cannot exceed unity.

Sufficieny:

An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter.

If $T = t(x_1, x_2, \ldots, x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2, \ldots, x_n$ of size n from the population with density $f(x, \theta)$ such that the conditional distribution of $x_1, x_2, \ldots, x_n$ given T_1 is independent of θ, then T is sufficient estimator for θ.

FACTORISATION THEOREM (NEYMANN):

The ~~nec~~ necessary and sufficient condition for a distribution to admit sufficient stastistic is provided by the 'factorization theorem' due to Neymann.

Statement:

$T = t(x)$ is sufficient for θ if and only if the joint density function L (say), of the sample values can be expressed in the form.

$$L = g_\theta\left[t(x)\right] \cdot h(x)$$

where (as indicated) $g_\theta\left[t(x)\right]$ depends on θ and x only through the value of $t(x)$ and $h(x)$ is independent of θ.

CRAMER - RAO INEQUALITY.

Definitions:-

If t is an unbiased estimator for $\gamma(\theta)$, a function of Parameter θ, then

$$var(t) \geq \frac{\left\{\frac{d}{d\theta}\cdot\gamma(\theta)\right\}^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} = \frac{\{\gamma'(\theta)\}^2}{I(\theta)}$$

where $I(\theta)$ is the information on θ, supplied by the sample.

In other words, Cramer-Rao inequality provides a lower bound $\{\gamma(\theta)\}^2/I\theta$, to the variance of an unbiased estimator of $\gamma(\theta)$.

Proof:

In Proving this result, we assume that there is only a single Parameter θ which is unknown. we also take the case of continuous r.v. The case of descrete random variables can be dealt with similarly on replacing the multiple integrals by appropriate multiple sums.

We further make the fg assumptions, which are known as the Regularity conditions for Cramer-Rao Inequality.

(1) The Parameter space Θ is a non-degenerate open interval on the real line $R'(-\infty,\infty)$.

(2) For almost all $x = (x_1, x_2, \ldots, x_n)$ and for all $\theta \in \Theta$, $\frac{\partial}{\partial\theta} L(x,\theta)$ exists, the exceptional set, if any is independent θ.

3) The range of integration is independent of the Parameter θ, so that $f(x,\theta)$ is differentiable under integral sign.

If range is not independent of θ and f is zero at the extremes of the range, i.e., $f(a,\theta) = 0 = f(b,\theta)$

then

$$\frac{\partial}{\partial\theta}\int_a^b f\,dx = \int_a^b \frac{\partial f}{\partial\theta}\,dx - f(a,\theta)\frac{\partial a}{\partial\theta} + f(b,\theta)\frac{\partial b}{\partial\theta}$$

$$\Rightarrow \frac{\partial}{\partial\theta}\int_a^b f\,dx = \int_a^b \frac{\partial f}{\partial\theta}\,dx, \text{ since } f(a,\theta) = 0 = f(b,\theta)$$

4) The conditions of uniform convergence of integrals are satisfied so that differentiation under the integral sign is valid.

5) $I(\theta) = E\left[\left\{\frac{\partial}{\partial\theta}\log L(x,\theta)\right\}^2\right]$, exists and is positive for all $\theta \in \Theta$

Let X be a r.v. following the p.d.f $f(x,\theta)$ and let L be the ~~ind~~ likelihood function of the random sample $(x_1, x_2, \ldots, x_n)$ from this Population. Then

$$L = L(x,\theta) = \prod_{i=1}^{n} f(x_i,\theta)$$

since L is the joint p.d.f of $(x_1, x_2, \ldots, x_n)$.

$$\int L(x,\theta)\,dx = 1$$

where $\int dx = \int\int\cdots\int dx_1\,dx_2\ldots dx_n$.

Differentiating with respect to θ and using regularity conditions given above, we get.

$$\int \frac{\partial}{\partial\theta} L\,dx = 0 \Rightarrow \int \left(\frac{\partial}{\partial\theta}\log L\right) L\,dx = 0 \Rightarrow E\left(\frac{\partial}{\partial\theta}\log L\right) = 0$$

Let $t = t(x_1, x_2, \ldots, x_n)$ be an unbiased estimator of $\gamma(\theta)$

such that

$$E(t) = \gamma(\theta) \Rightarrow \int t \cdot L\,dx = \gamma(\theta)$$

Differentiating w.r to θ, we get

$$\int t \cdot \frac{\partial L}{\partial\theta}\,dx = \gamma'(\theta) \Rightarrow \int t\left(\frac{\partial}{\partial\theta}\log L\right) L\,dx = \gamma'(\theta)$$

$$\Rightarrow E\left(t \cdot \frac{\partial}{\partial\theta}\log L\right) = \gamma'(\theta)$$

$$\mathrm{cov}\left(t, \frac{\partial}{\partial\theta}\log L\right) = E\left(t \cdot \frac{\partial}{\partial\theta}\log L\right) - E(t) \cdot E\left(\frac{\partial}{\partial\theta}\log L\right)$$

$$= \gamma'(\theta)$$

We have :

$$\{r(x,y)\}^2 \le 1 \Rightarrow \{\mathrm{cov}(x,y)\}^2 \le \mathrm{var}(x) \cdot \mathrm{var}(y)$$

$$\therefore \left\{\mathrm{cov}\left(t, \frac{\partial}{\partial\theta}\log L\right)\right\}^2 \le \mathrm{var}\,t \cdot \mathrm{var}\left(\frac{\partial}{\partial\theta}\log L\right)$$

$$\Rightarrow \{r'(\theta)\}^2 \le \mathrm{var}\,t \left[E\left(\frac{\partial}{\partial\theta}\log L\right)^2 - \left\{E\left(\frac{\partial}{\partial\theta}\log L\right)\right\}^2\right]$$

$$\Rightarrow \{r'(\theta)\}^2 \le \mathrm{var}\,t \cdot E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}$$

$$\Rightarrow \mathrm{var}(t) \ge \frac{\{\gamma'(\theta)\}^2}{E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}}$$

which is Cramer-Rao inequality.

Theorem 17.17 [INVARIANCE PROPERTY OF MLE]. If T is the MLE of θ and $\psi(\theta)$ is one function of θ, then $\psi(T)$ is the MLE of $\psi(\theta)$

Example 17.31. In random sampling from normal population $N(\mu, \sigma^2)$, find the maximum likelihood estimates for.

(i) μ when σ^2 is known, (ii) σ^2 when μ is known, and iii) the simultaneous estimation of μ and σ^2

Soln:

$$X \sim N(\mu, \sigma^2), \text{ then}$$

$$L = \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}\right] = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^{n}(x_i-\mu)^2 / 2\sigma^2\right\}$$

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

case(i): when σ^2 is known, the likelihood equation for estimating μ is:

$$\frac{\partial}{\partial \mu}\log L = 0 \Rightarrow -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Hence M.L.E for μ is the sample mean $\bar{x}$

case(ii) when μ is known, the likelihood equation for estimating σ^2 is:

$$\frac{\partial}{\partial \sigma^2}\log L = 0 \Rightarrow -\frac{n}{2} \times \frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow n - \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

case(iii) the likelihood equations for simultaneous estimation of μ and σ^2 are:

$\frac{\partial}{\partial \mu}\log L = 0$ and $\frac{\partial}{\partial \sigma^2}\log L = 0$, thus giving $\hat{\mu} = \bar{x}$.

and

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s^2$, the sample variance.

Important Note:- It may be pointed out here that though

$$E(\hat{\mu}) = E(\bar{x}) = \mu,\ E(\hat{\sigma}^2) = E(s^2) \neq \sigma^2$$

Hence the maximum likelihood estimators (M.L.E s) need not necessarily be unbiased. Another illustration is given.

Remark:

Since M.L.E is the most efficient, we conclude that in sampling from a normal population, the sample mean $\bar{x}$ is the most efficient estimator of the Population mean μ.

Example 17.33.

a) Find the maximum likelihood estimate for the Parameter λ of a Poisson distribution on the basis of a sample of size n. Also find its variance.

b) Show that the sample mean $\bar{x}$, is sufficient for estimating the parameter λ of the poisson distribution.

Soln:

The probability function of the poisson distribution with Parameter λ is given by.

$$P(X=x) = f(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \ ;\ x = 0, 1, 2, \ldots$$

Likelihood function of the random sample $x_1, x_2, \ldots x_n$ of n observations from this population is:

$$L = \prod_{i=1}^{n} f(x_i, \lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{x_1!\, x_2! \ldots x_n!}$$

$$\therefore \log L = -n\lambda + n\bar{x}\log\lambda - \sum_{i=1}^{n}\log(x_i!)$$

The likelihood equation for estimating λ is:

$$\frac{\partial}{\partial\lambda}\log L = 0 \Rightarrow -n + \frac{n\bar{x}}{\lambda} = 0 \Rightarrow \lambda = \bar{x}$$

Thus the M.L.E for λ is the sample mean $\bar{x}$. The variance of estimate is given by:

$$\frac{1}{V(\hat{\lambda})} = E\left\{-\frac{\partial^2}{\partial\lambda^2}(\log L)\right\}$$

$$= E\left\{-\frac{\partial}{\partial\lambda}\left(-n+\frac{n\bar{x}}{\lambda}\right)\right\} = E\left\{-\left(\frac{-n\bar{x}}{\lambda^2}\right)\right\}$$

$$= \frac{n}{\lambda^2}E(\bar{x}) = n/\lambda$$

$$V(\hat{\lambda}) = \lambda/n$$

b) For the Poisson distribution with parameter λ, we have

$$\frac{\partial}{\partial\lambda}\log L = -n+\frac{n\bar{x}}{\lambda} = n\left(\frac{\bar{x}}{\lambda}-1\right) = \psi(\bar{x},\lambda),$$ a function of $\bar{x}$ and λ only.

Hence (c.f. Remark to theorem 17.15). $\bar{x}$ is sufficient for estimating λ.

Example 17.38

State as precisely as possible the properties of the M.L.E obtain the M.L.E.s of α and β for a random sample from the exponential population.

$f(x;\alpha,\beta) = y_0 e^{-\beta(x-d)}$, $\alpha \le x \le \infty$, $\beta > 0$ and y_0 being a constant.

Soln: Here first of all we shall determine the constant y_0 from the consideration that the total area under a probability curve is unity.

$$\therefore y_0 \int_{\alpha}^{\infty} \exp\left[-\beta(x-\alpha)\right]dx = 1$$

$$\Rightarrow y_0 \left|\frac{e^{-\beta(x-\alpha)}}{-\beta}\right|_{\alpha}^{\infty} = 1$$

$$\Rightarrow \frac{-y_0}{\beta}(0,-1) = 1 \Rightarrow y_0 = \beta$$

$$f(x;\alpha,\beta) = \beta e^{-\beta(x-\alpha)}, \quad \alpha \le x < \infty$$

If $x_1, x_2, \ldots, x_n$ is a random sample of n observations from this population, then

$$L = \prod_{i=1}^{n} f(x_i;\alpha,\beta) = \beta^n \exp\left\{-\beta \sum_{i=1}^{n}(x_i-\alpha)\right\}$$
$$= \beta^n \exp\left[-n\beta(\bar{x}-\alpha)\right]$$

$$\log L = n\log\beta - n\beta(\bar{x}-\alpha) \rightarrow ①$$

The likelihood equations for estimating α and β give

$$\frac{\partial}{\partial\alpha}\log L = 0 = n\beta \rightarrow ②$$

and $$\frac{\partial}{\partial\beta}\log L = 0 = n/\beta - n(\bar{x}-\alpha) \rightarrow ③$$

Equation ② gives $\beta=0$, which is obviously inadmissible and this on substitution in ③ gives $\alpha=\infty$, a nugatory result. Thus the likelihood equations fail to give us valid estimates of α and β and we try to locate M.L.E.S for α and β by maximising L directly.

L is maximum $\Rightarrow \log L$ is maximum

From ①, $\log L$ is maximum (for any value of β), if $(\bar{x}-\alpha)$ is minimum, which is so if α is maximum. If $x(1), x(2), \ldots, x(n)$ is ordered sample from this Population then $\alpha \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} < \infty$, so that the maximum value of α consistent with the sample is $x_{(1)}$, the smallest sample observation i.e, $\hat{\alpha} = x_{(1)}$.

Consequently (3) gives $1/\beta = \bar{x} - \hat{\alpha} = \bar{x} - x_{(1)} \Rightarrow \hat{\beta} = \dfrac{1}{\bar{x}-x_{(1)}}$

Hence M.L.E.S for α and β are given by:

$$\hat{\alpha} = x_{(1)} \quad \text{and} \quad \hat{\beta} = \frac{1}{\bar{x}-x_{(1)}}$$

Example 15.44

For the double poisson distribution

$$P(x) = P(X=x) = \frac{1}{2}\cdot\frac{e^{-m_1}\cdot m_1^x}{x!} + \frac{1}{2}\cdot\frac{e^{-m_2}\cdot m_2^x}{x!}\ ;\ x = 0,1,2,\ldots$$

show that the estimates for m_1 and m_2 by the method of moments are: $\mu_1 \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$

Soln: we have

$$\mu_1' = \sum_{x=0}^{\infty} x\cdot p(x) = \frac{1}{2}\sum_{x=0}^{\infty} x\cdot\frac{e^{-m_1}\cdot m_1^x}{x!} + \frac{1}{2}\sum_{x=0}^{\infty} x\cdot\frac{e^{-m_2}\cdot m_2^x}{x!}$$

$$= \frac{1}{2}m_1 + \frac{1}{2}m_2 \rightarrow (1)$$

[since the first and second summations are the means of Poisson distributions with parameters m_1 and m_2 respectively)

$$\mu_2' = \sum_{x=0}^{\infty} x^2 p(x) = \frac{1}{2}\left\{\sum_{x=0}^{\infty} x^2\cdot\left(\frac{e^{-m_1}m_1^x}{x!}\right) + \sum_{x=0}^{\infty} x^2\cdot\left(\frac{e^{-m_2}m_2^x}{x!}\right)\right\}$$

$$= \frac{1}{2}\left\{(m_1^2+m_1) + (m_2^2+m_2)\right\}$$

$$= \tfrac{1}{2}\left\{(m_1+m_2) + (m_2^2+m_2)\right\}$$

$$\mu_2' = \frac{1}{2}\left\{(m_1+m_2) + (m_1^2+m_2^2)\right\} \rightarrow (2)$$

$$= \tfrac{1}{2}\left\{2\mu_1' + m_1^2 + (2\mu_1' - m_1)^2\right\}$$

$$= \tfrac{1}{2}\left(2\mu_1' + m_1^2 + 4\mu_1'^2 + m_1^2 - 4m_1\mu_1'\right)$$

$$\Rightarrow \mu_2' = \mu_1' + m_1^2 + 2\mu_1'^2 - 2\mu_1' m_1$$

$$\Rightarrow m_1^2 - 2m_1\mu_1' + (2\mu_1'^2 + \mu_1' - \mu_2') = 0$$

$$\therefore \hat{m}_1 = \frac{2\mu_1' \pm \sqrt{4\mu_1'^2 - 4(2\mu_1'^2 + \mu_1' - \mu_2')}}{2}$$

Example 15.44

For the double poisson distribution

$$P(x) = P(X=x) = \frac{1}{2}\cdot\frac{e^{-m_1}\cdot m_1^x}{x!} + \frac{1}{2}\cdot\frac{e^{-m_2}\cdot m_2^x}{x!}\ ;\ x = 0, 1, 2, \ldots$$

Show that the estimates for m_1 and m_2 by the method of moments are: $\mu_1' \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$

Soln: We have

$$\mu_1' = \sum_{x=0}^{\infty} x\cdot p(x) = \frac{1}{2}\sum_{x=0}^{\infty} x\cdot\frac{e^{-m_1}\cdot m_1^x}{x!} + \frac{1}{2}\sum_{x=0}^{\infty} x\cdot\frac{e^{-m_2}\cdot m_2^x}{x!}$$

$$= \frac{1}{2}m_1 + \frac{1}{2}m_2 \rightarrow (1)$$

[Since the first and second summations are the means of Poisson distributions with parameters m_1 and m_2 respectively)

$$\mu_2' = \sum_{x=0}^{\infty} x^2 p(x) = \frac{1}{2}\left\{\sum_{x=0}^{\infty} x^2\cdot\left(\frac{e^{-m_1} m_1^x}{x!}\right) + \sum_{x=0}^{\infty} x^2\cdot\left(\frac{e^{-m_2} m_2^x}{x!}\right)\right\}$$

$$= \frac{1}{2}\left\{(m_1^2 + m_1) + (m_2^2 + m_2)\right\}$$

$$= \frac{1}{2}\left\{(m_1 + m_2) + (m_2^2 + m_2)\right\}$$

$$\mu_2' = \frac{1}{2}\left\{(m_1 + m_2) + (m_1^2 + m_2^2)\right\} \rightarrow (2)$$

$$= \frac{1}{2}\left\{2\mu_1' + m_1^2 + (2\mu_1' - m_1)^2\right\}$$

$$= \frac{1}{2}\left(2\mu_1' + m_1^2 + 4\mu_1'^2 + m_1^2 - 4m_1\mu_1'\right)$$

$$\Rightarrow \mu_2' = \mu_1' + m_1^2 + 2\mu_1'^2 - 2\mu_1' m_1$$

$$\Rightarrow m_1^2 - 2m_1\mu_1' + (2\mu_1'^2 + \mu_1' - \mu_2') = 0.$$

$$\therefore \hat{m}_1 = \frac{2\mu_1' \pm \sqrt{4\mu_1'^2 - 4(2\mu_1'^2 + \mu_1' - \mu_2')}}{2}$$

$$= \mu_1' \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$$

similarly on substituting for m_1 in terms of m_2 from(1) in (2) we get.

$$m_2^2 - 2m_2\mu_1' + (2\mu_1'^2 + \mu_1' - \mu_2') = 0$$

solving for m_2, we get

$$\hat{m}_2 = \mu_1' \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$$